

# Wyoming's Statewide Assessment System: Recommendations from the Wyoming Assessment Task Force

---

Compiled By

Joseph Martineau, Ph.D. and Scott Marion, Ph.D.  
National Center for the Improvement in Educational Assessment

Draft September 4, 2015

October 2015

## TABLE OF CONTENTS

Executive Summary.....	iii
Chapter 1: Appropriate Characteristics and Uses of Assessment.....	4
Types of Assessments and Appropriate Uses .....	4
Chapter 2: Desired Characteristics and Uses .....	10
High-Priority Characteristics and Uses .....	10
Chapter 3: Intended Outcomes .....	12
Chapter 4: Narrative Recommendations for a Comprehensive Assessment System .....	15
Context.....	15
Proposed Wyoming Assessment System .....	15
Supports and Conditions.....	19
Ensuring a Successful Transition .....	20
Chapter 5: Detailed Recommendations.....	22
State Summative Assessment.....	22
Interim Assessment.....	28
District Assessment System.....	28
Classroom Formative Assessment.....	29
Chapter 6: Recommendations for Policy Coherence .....	30
References.....	31
Appendix A: Vignettes of Formative Assessment Practices .....	32
Appendix B: One-Page Summary of Formative, Interim, and Summative Assessment.....	34
Appendix C: Detailed Highest Priority Uses and Characteristics.....	35

## EXECUTIVE SUMMARY

Placeholder.

DRAFT

## CHAPTER 1: APPROPRIATE CHARACTERISTICS AND USES OF ASSESSMENT

In 2015, the Wyoming Legislature passed Enrolled Act 87, authorizing the State Board of Education to evaluate Wyoming's current state assessment system and the creation of the Wyoming Assessment Task Force. Specifically, Section 6 of the act authorizes:

*The state board shall assemble a task force to assist with the assessment review and evaluation. The task force shall be comprised of representatives of small and large school districts and schools from all geographic regions of the state and shall at minimum include representatives from district and school administration, school district assessment and curriculum program administrators, elementary and secondary school teachers, school district board members, state higher education representatives, member of the Wyoming business community and parents of children enrolled in Wyoming public schools.*

The twenty-four task force members<sup>1</sup> met seven times between June 1 and October 1, 2015. Three of these meetings were held in person, one of which was for two full days, and the remaining four meetings were held as webinars. This report presents the results of the task force deliberations. Before moving to the discussion of the task force recommendations, we first present in this chapter some critical definitions and background assessment information on the appropriate characteristics and use of assessment.

We begin by defining two broad categories of assessment use: (1) *high-stakes accountability uses* and (2) *lower-stakes instructional uses*. Stakes may be high for students, teachers or administrators, or schools and districts. For students, test scores may be used for making high-stakes decisions regarding grades, grade promotion, tracking, graduation, admission to postsecondary education or training, and scholarships. For educators, student test scores may formally or informally factor into periodic evaluations used to inform important employment decisions in classrooms, departments, schools and districts. In addition, students, teachers and administrators are affected by high-stakes uses of test scores in school and district accountability: identification as a school or district in need of intervention often leads to involuntary interventions intended to correct poor outcomes.

*Lower-stakes instructional* uses of test scores for teachers and administrators include informing moment-to-moment instruction; self-evaluation in teaching a unit and adjusting subsequent plans accordingly, evaluating one's own instructional effectiveness; and evaluating the success of a curriculum, program, or intervention.

As described above, within the *high stakes accountability* and *lower stakes formative* categories of use, there are many potential uses. The multiple appropriate uses of the various types of assessment introduced below may fall into both broad categories.

### Types of Assessments and Appropriate Uses

While there are several possible categorizations of assessment by type, this chapter reviews only one particularly relevant to the work of the Task Force: the distinction among *formative*, *interim*, and

---

<sup>1</sup> There were 26 original members, but two members resigned during the course of the project due to other commitments.

summative assessment<sup>2</sup>. This section provides definitions of the three types of assessment and outlines the appropriate uses of data gathered from them. Appropriate uses are underlined for emphasis. These definitions are critical to understanding what each type of assessment can and cannot do.

We define formative, interim, and summative assessment below. Appendix B provides an at-a-glance summary of the typical characteristics, appropriate uses, and examples of each type of assessment.

## Formative Assessment

Formative assessment has also been called formative instruction. The purpose of formative assessment is to evaluate student understanding against key learning targets, provide targeted feedback to students, and adjust instruction on a moment-to-moment basis.

In 2006, the Council of Chief State School Officers (CCSSO) and experts on formative assessment developed a widely cited definition (Wiley, 2008):

*Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievements of intended instructional outcomes (p. 3).*

In addition, Wiley (paraphrased from p. 3) lists five critical attributes of formative assessment:

1. They are based on clear articulations of learning goals as steps toward an ultimate desirable outcome.
2. Learning goals and the criteria for success are clearly identified and communicated to students in language they can understand.
3. Students are frequently provided with feedback directly linked to the learning goals and criteria for success.
4. Students engage in self- and peer-assessment against the criteria for success.
5. Students and teachers jointly own (collaborate on) monitoring student progress over time.

While the practice of formative assessment in general embodies these five attributes, not every example of formative assessment incorporates every attributes. The definition given above and five critical attributes are based on research linking such practices to student learning gains. The core of the formative assessment process is that it takes place during instruction (i.e., “in the moment”) and under full control of the teacher to support student learning while it is developing. This is done through diagnosing on a very frequent basis where students are in their progress toward learning goals, where gaps in knowledge and skill exist, and how to help students close those gaps.

Another important conclusion one can draw from the definitions and critical attributes is that formative assessment is embedded within instruction. Instruction does not stop when teachers engage in formative assessment. Formative assessment covers fine-grained learning targets that are often the focus of a single unit of instruction.

---

<sup>2</sup> In defining formative, interim, and summative assessment, this section borrows from three sources (Perie, Marion, Gong, & Wurtzel, 2007; Michigan Department of Education, 2013; Wiley, 2008).

The definition and critical attributes make clear that formative assessment is not a product, but a process tailored to the details of ongoing instruction to individual students. Effective formative assessment practices occur very frequently, covering very small units of instruction (such as part of a class period). If tasks are presented, they may vary for students depending on where they are in their learning. However, formative assessment processes often occur during regular and targeted questioning of students in small or large groups, observing students as they work in groups and/or engage in tasks. Formative assessment practices may be facilitated using certain technology and related tools. There is a strong view among some scholars that because formative assessment is tailored to the specific context of the classroom and to individual students that results cannot be meaningfully aggregated or compared. Many of these scholars question whether the observations from formative assessment should even be scored.

Another conclusion that may be drawn from this definition of formative assessment is the critical importance of providing frequent feedback to individual students. For each student such feedback develops the ability to continuously monitor the quality of their own work against a clear learning target. It is this feedback to students that is the most crucial part of the formative assessment process<sup>3</sup>.

The nature of formative assessment implies that the frequently used term *common formative assessment* is an inappropriate use of the term. Effective formative assessment is tailored to a specific instructional plan and a specific group of students at defined points in their attainment of learning targets. The critical characteristics of formative assessment practices should be common across all teachers, and tools teachers use to implement formative assessment may be common across many teachers, but formative assessment is too tailored to a unique classroom to be common.

Data gathered through formative assessment have limited to no use for evaluation or accountability purposes such as student grades, educator accountability, school/district accountability, or even public reporting that could allow for inappropriate comparisons. There are at least four reasons for this: (1) if carried out appropriately, the data gathered from one unit to the next, one teacher to the next, one moment to the next, and one student to the next will not be comparable; (2) students will be unlikely to participate as fully, openly, and honestly in the process if they know they are being evaluated by their teachers or peers on the basis of their responses; (3) for the same reasons, educators will be unlikely to participate as fully, openly, and honestly in the process; and (4) the nature of the formative assessment process is likely to shift in such a way that it can no longer optimally inform instruction.

These implications create a distinct difference from summative and interim assessment (described below), which are intended to assess student achievement after an extended period of learning. Simply giving students an assessment in the classroom does not mean that the assessment is formative. Use of assessment evidence in a formative manner requires teachers to achieve insight into individual student learning in relation to learning targets, to provide effective feedback to students about those insights, and to make instructional decisions based on those insights. During the formative assessment process, feedback to students and student involvement is essential. Teachers seek ways to involve the student in “thinking about their thinking” (metacognition) to use learning evidence to close the gap and get closer to the intended learning target.

---

<sup>3</sup> See Sadler (1989).

Because there is considerable confusion about what formative assessment is, four vignettes<sup>4</sup> describing formative assessment in action are given in Appendix A to clarify the meaning using concrete ideas. The first two vignettes are also presented in condensed form in the one-page summary of formative, interim, and summative assessment in Appendix B.

## Summative Assessment

Summative assessments are generally infrequent (e.g., administered only once to any given student) and cover major units of instruction such as semesters, courses, credits, or grade levels. They are typically given at the end of a defined period to evaluate students' performance against a set of learning targets for the instructional period. The prototypical assessment conjured by the term "summative assessments" is given in a standardized manner statewide (but can also be given nationally or districtwide) and is typically used for accountability or to otherwise inform policy. Such summative assessments are typically the least flexible of the various assessment types. Summative assessments are also used for testing out of a course, diploma endorsement, graduation, high school equivalency, and college entrance. Appropriate uses of such standardized summative assessments include school accountability, district accountability, curriculum evaluation, program evaluation, and informing policy-makers in high-level decision-making. Depending on their alignment to classroom instruction and the timing of the administration and results, they may also be appropriate for grading.

Less standardized, but no less summative, assessments are also found in the majority of middle- and high-school classrooms. Such assessments are typically completed near the end of a semester, credit, course, or grade level. Common examples are broad exams or projects intended to give a summary of student achievement of marking period objectives, and figure heavily in student grading. Such assessments tend to be labeled "mid-terms," "final projects," "final papers," or "final exams" in middle and high school grades. Elementary school classrooms also have similar summative assessments but these do not have a consistent label in elementary grades. Classroom summative assessments may be created by individual teachers or by staff from one or more schools or districts working together.

Summative assessments tend to require a pause in instruction for test administration. They may be controlled by a single teacher (for assessments unique to the classroom), groups of teachers working together, a school (e.g., for all sections of a given course or credit), a district (to standardize across schools), a group of districts working together, a state, a group of states, or a test vendor. The level at which test results are comparable depends on who controls the assessment. They may be comparable within a classroom, across a few classrooms, within a school, within a district, across a few districts, within a state, or across multiple states.

Appropriate uses of such summative assessments include student grading in the specific courses for which they were developed. If designed well, they can also be used to adjust curriculum, programming, and instruction the next time the large unit of instruction is taught; and to serve as a post-test measure of student learning. If the assessments are well-designed and a carefully- and well-defined set of rules is in place for appropriate administration, scoring, and use of results they may also be reasonably used for accountability.

---

<sup>4</sup> Informed by Wiley (2008).

## Interim Assessment

Many periodic standardized assessment products currently in use that are marketed (or otherwise labeled) as “formative,” “benchmark,” “diagnostic,” or “predictive” actually belong in the interim assessment category. They are neither formative (they do not facilitate moment-to-moment targeted analysis of student learning, frequent feedback to students and teachers, or timely adjustment of instruction) nor summative (they are not intended to provide a broad summary of achievement of course- or grade-level learning objectives tied to specific state content standards). In contrast to formative assessment

Many interim assessments are commercial products and rely on fairly standardized administration procedures that provide information relative to a specific set of learning targets—although not always tied to specific state content standards—and are designed to inform decisions at the classroom, school, and/or district level. In some cases, interim assessments may be controlled at the classroom level to provide information for the teacher, but unlike formative assessment, the results of interim assessments can be meaningfully aggregated and reported at a broader level. However, the adoption and timing of such interim assessments are likely to be controlled by the school district. The content and format of interim assessments is also very likely to be controlled by the test developer. Therefore, these assessments are considerably less instructionally-relevant than formative assessments in that decisions at the classroom level tend to be *ex post facto* regarding post-unit remediation needs and adjustment of instruction the next time the unit is taught.

Common assessments developed by a school or district for the purpose of measuring student achievement multiple times throughout a year may be considered interim assessments. These may include common mid-term exams and other periodic assessments such as quarterly assessments.

Standardized interim assessments may be appropriate for a variety of uses, including predicting a student’s likelihood of success on a large-scale summative assessment, evaluating a particular educational program or pedagogy, identifying potential gaps in a student’s learning after a limited period of instruction has been completed, or measuring student learning over time.

There are three other types of interim assessments currently in use beyond the “backward looking” interim assessments described above. All are “forward-looking.” One useful but less widely used type is a pre-test given before a unit of instruction to gain information about what students already know in order to adjust plans for instruction before beginning the unit (teachers may do these pre-instruction checks on a more frequent, formative basis). Such forward-looking assessments may be composed of pre-requisite content or the same content as the end-of-unit assessment. A second type of forward-looking assessment is a placement exam used to personalize course-taking according to existing knowledge and skills. Finally, a third type of forward-looking assessment is intended to predict how a student will do on a summative assessment before completing the full unit of instruction. The usefulness of this type of interim assessment is debatable in that it is unlikely to provide much instructionally relevant information and there is often other information available to determine who is likely to need help succeeding on the end of year summative assessment.

## A Note on Classroom Assessment and Accountability



If considerable resources are provided to support classroom-level formative, interim, and summative assessment, there may be a reasonable question as to whether funds are being invested wisely. One temptation may be to hold educators, schools, and/or districts accountable for results on classroom assessments, but such uses are inappropriate for formative and interim assessment, and great care is needed when using classroom summative assessments in such ways. Rather than holding schools and/or teachers accountable for student data gathered from classroom interim and formative assessment, the investment could be evaluated instead by:

- Monitoring the *quality* of formative, interim, and summative classroom assessment practices *rather than outcomes* based on those assessments in such a way that encourages collaboration.
- Requiring teachers and administrators to attend high-quality professional development (PD) on best practices in classroom assessment.
- Monitoring the *degree and quality of administrator support* for teachers to collaborate and improve their formative, interim, and summative classroom assessment practices *rather than outcomes* based on those assessments.

If student *data* from formative or interim classroom assessment are used for educator or school accountability, implementation is likely to be corrupted, and beneficial instructional effects of the investment are likely to be lost.

## CHAPTER 2: DESIRED CHARACTERISTICS AND USES

With the background of appropriate characteristics and uses of assessment from Chapter 1, it is possible to have a coherent presentation of the desired characteristics, uses, and outcomes of assessment as developed by the Task Force.

### High-Priority Characteristics and Uses

The Task Force considered that assessment design is always a case of optimization under constraints<sup>5</sup>. In other words, there may be many desirable purposes, uses, and goals for assessment. However, they may be in conflict. Any given assessment can serve only a limited number of purposes well. Finally, assessments always have some type of restrictions (e.g., legislative requirements, time, cost, etc...) that must be weighed in finalizing recommendations.

Task Force members initially were asked to ignore constraints, and identify their desired purposes and goals for assessment and their desired uses of assessment data. Subgroups of Task Force members noted their highest priority uses, and then reviewed the work of other subgroups, asking clarifying questions. After each subgroup's highest priority uses and purposes were reviewed, each individual panelist identified their three highest priorities. The full task force then discussed possible patterns emerging from the activity.

In general, Task Force members desire a Wyoming assessment (system) that is capable of serving the following broad purposes:

- Provide instructionally-useful information to teachers and students (with appropriate grain-size and timely reporting)
- Provide clear and accurate information to parents and students regarding students' achievement of and progress toward key outcomes, such as progress toward meeting grade-level standards and progress toward readiness for post-secondary education and/or career training
- Provide meaningful information to support evaluation and enhancement of curriculum and programs
- Provide information to appropriately support federal and state accountability determinations

Top priority uses and characteristics that were similar were consolidated. In consolidating, important differences in each contributing uses/characteristics were incorporated into the consolidated description. Appendix B provides more detailed information regarding this prioritization activity.

An important outcome of this activity is that no single type of assessment (formative, interim, or summative) is applicable to all of the high-priority desired uses and characteristics. In fact, formative assessment is uniquely able to support two uses/characteristics and summative assessment is uniquely able to support three uses/characteristics. The same is true for level of assessment: classroom-level and state-level assessment are each uniquely able so support three uses/characteristics.

---

<sup>5</sup> See Braun (in press).

These outcomes of the Task Force's work indicate that in order to accomplish the full set of uses and characteristics, **a system of assessments** would be required that span the range of assessment type (formative, interim, and summative) and assessment level (classroom, district, and state). This can be accomplished by combining state and local assessments in a way that they create a coherent system that eliminates unnecessary assessment and provides a consistent picture with complementary characteristics and uses.

### **A Statewide Summative Assessment or an Assessment System?**

As mentioned above, a single assessment is incapable of meeting the various high-priority characteristics and uses identified by the Task Force. In order to do so, all three types of assessment may be necessary. However, in the same way that a pile of bricks does not make a house, a collection of assessments at the classroom, school, district, and state level is not necessarily a coherent assessment system capable of meeting multiple intended uses<sup>6</sup>.

It is clear that the Task Force desires to respect local control, maintain the autonomy of individual educators, and provide educators appropriate professional development and ongoing support. Designing a comprehensive assessment system within statutory constraints that also meets the desires listed above is difficult and complex, but not impossible. **Based on these considerable difficulties and complexities, the Task Force was faced with a decision: Recommend a single statewide summative assessment to fulfill statutory requirements or a comprehensive assessment system.**

The Task Force first voted to explore the possibility of a comprehensive assessment system (with a few members expressing reluctance and reserving judgment). After further discussion in later meetings, **Task Force members unanimously voted to make recommendations for a comprehensive assessment system.** As a prelude to the specific recommendations, Task Force members identified issues with the existing state, interim, and district assessments that should be addressed in developing recommendations. They also developed intended outcomes based on those issues. Those issues and intended outcomes are presented in Chapter 3. A narrative summary of the Task Force recommendations for addressing those issues and achieving the intended outcomes is provided in Chapter 4. Detailed recommendations to assist in developing a request for proposal (RFP) and in evaluating bids are provided in Chapter 5. Changes to policy necessary to allow for implementation are presented in Chapter 6.

---

<sup>6</sup> See Coladarci (2002).

## CHAPTER 3: INTENDED OUTCOMES

In developing recommendations for a new state summative assessment, the Task Force deliberated on issues it intended to address in developing its recommendations in three areas: state summative assessment, interim assessments, and district assessment systems. The issues identified by the Task Force include the following:

### Issues with State Summative Assessment

#### *Timing and Stability*

- The state summative assessment is administered too early in the year to reflect a full year of instruction, and on the flip side results sometimes come too late for use in school improvement activities such as program and curriculum evaluation. The assessment needs to be administered later in the year *and* results need to be returned in time for use in school improvement.
- The use of state test scores for school improvement activities is tenuous because the test or the cut scores on the test change too often. The state assessment needs to remain stable for many years to allow for analysis of policies, programming, and curriculum over time.
- Comparing results from Wyoming state assessment to other states is not possible because the assessment is unique to Wyoming. It is important that Wyoming be able to compare its results with other states with similar content standards to inform state and local policy.

#### *Test Quality*

- The quality and usefulness of student achievement and growth reports needs to be improved, given the high-stakes use of state test results. It is important that the state assessment include high-level tasks representative of the kind of teaching we expect from Wyoming educators.
- It is important for the test to represent both the depth and breadth of the Wyoming state content standards. Multiple-choice-only tests are inadequate in that they signal that Wyoming puts a priority on easy-to-measure knowledge and skills.

#### *Concerns about Appropriate Use*

- Educators need adequate professional development in appropriate uses of state assessment data and appropriate preparation for success on the assessment. Teachers need confidence that they can appropriately use state assessment data to improve their own practice.
- Educators need adequate professional development in appropriate uses of state assessment data and appropriate preparation for success on the assessment. Teachers need confidence that they can appropriately use state assessment data to improve their own practice.
- Current use of ACT goes beyond what is appropriate. The ACT is a college entrance examination that is built to measure ACT's college readiness standards. It was not developed to measure the Wyoming state content standards. As such, it is inappropriate to use the ACT as the sole accountability assessment in high school. The use of college entrance assessment

scores should be limited to the use for which it has been validated: predicting college success.

- The use of ACT as the sole high school accountability assessment has resulted in confusion about what the high school learning targets are: the official Wyoming state standards or the ACT college readiness standards? Wyoming high school educators need the high school learning targets to be clear in order to appropriately align their instruction to one set of learning targets.

### **Issues with Interim Assessment**

The Task Force identified incoherence between the existing state assessment and the various interim assessment currently in use as an issue. It is important for the state and interim assessments to provide consistent information about individual students and groups of students to assure that difference seen in the results are not simply artifacts of differences between the tests in terms of format, quality, and content coverage.

### **Issues with District Assessment Systems**

While Wyoming districts have been responsible for developing local assessment systems for a long time, there has been little review of the technical quality of such assessment systems. The Task Force recognized the need for improving the quality of district assessments to increase their usefulness in informing local decisions and for documenting student learning of the basket of goods. The following three general issues were identified:

- Varying levels of coherence of district assessment systems with the state assessment and with interim assessments, leading to confusion in conclusions drawn from the various assessments.
- Varying degrees of quality of district assessment systems.
- Inadequate local capacity to develop and validate high-quality local assessment systems.
- Inadequate evaluation and quality control of local assessment systems.

### **Intended Outcomes of a Comprehensive Assessment System**

Based on desired characteristics and uses of assessment and on issues identified above, the Task Force developed the following broad intended outcomes of a new Wyoming Comprehensive Assessment System:

- Assist students (and their parents) to become more engaged in their own education through a greater knowledge of their strengths and weaknesses and their current (and likely future) achievement of desirable outcomes by providing daily feedback from formative assessment and periodic evaluative data from interim and summative assessment.
- Provide teachers and administrators with timely data on individual students' strengths and weaknesses, and their current and likely future achievement of desirable outcomes.
- Improve statewide achievement and growth rates over previous Wyoming achievement and other states through

1. Providing high-quality data, stable across many years, to high-level policymakers to inform changes in educational policy;
  2. Holding schools and districts appropriately accountable for student outcomes; and
  3. Providing valid data to local administrators in order to adjust programs and curriculum to target areas of weakness.
- Improve day-to-day integration of assessment with instruction by encouraging both teacher-level collaboration and material administrative support for initial and ongoing professional development and collaboration at the state, district, and school levels.
  - Achieve alignment of curriculum, instruction, and assessment with the officially adopted Wyoming state standards in every district to ensure that every Wyoming student is provided a high-quality opportunity to learn the “basket of goods.”
  - Achieve stability and coherence of local and state assessments to allow for a focus on improving instruction rather than a focus on adapting to new assessments.
  - Achieve stability of local and state assessments to allow for a single-minded focus on improving instruction rather than adapting to new assessments.
  - Achieve coherence of local, interim, and state assessments.
  - Improve the quality of district assessment systems.
  - By ending state accountability testing in grade 10:
    1. Avoid adding to grade 11 Advanced Placement (AP), International Baccalaureate (IB), and college entrance testing.
    2. Encourage universal development and use of individualized pathways through a Career & Technical Education (CTE) program and/or college preparation program.
    3. Strengthen ties to community colleges and universities to smooth students’ transitions from high-school to postsecondary education and/or training and to provide students with more options for valuable certificates at the time of graduation
    4. Improve equity in options available to all high-school students regardless of location
  - Prioritize the Wyoming state content standards in a transparent way so that educators clearly know what knowledge and skills will be included on the test and that the complete set of test-eligible content is feasible to teach in the allotted instructional time.
  - Expand beyond multiple choice to include other types of tasks on the state assessment more suited to measuring high-level knowledge and skills.
  - Convey to all Wyoming education stakeholders that high-quality writing is a valuable skill that must be effectively taught and learned in Wyoming public schools.

Chapter 4 provides an overview of the system recommended by the Task Force. Chapter 5 provides detailed recommendations. Chapters 4 and 5 are presented separately because it is difficult to get a coherent picture of what the assessment system would look like from the various detailed recommendations.

## **CHAPTER 4: NARRATIVE RECOMMENDATIONS FOR A COMPREHENSIVE ASSESSMENT SYSTEM**

Wyoming stakeholders have determined that they want an assessment system that will serve multiple purposes, designed to address several uses, including documenting Wyoming student academic achievement and growth rates as well as supporting local instructional and program evaluation needs. In order to achieve this goal, a thoughtfully-designed and coherent system of state, local, and classroom assessment is required.

The design of such a system is intended to create appropriate and high-quality data from all levels of the education system to support a variety of purposes. It is also intended to minimize or eliminate redundant assessments while maximizing coherence among the results from a variety of assessments. However, recognizing that even comprehensive assessment systems cannot serve every desirable purpose well, the stakeholders prioritized the following broad purposes for the “new” Wyoming Assessment System:

- Producing instructionally-useful information for teachers and students (with appropriate grain-size and timely reporting),
- Providing clear and accurate information to parents and students regarding students’ achievement of and progress toward key outcomes,
- Producing meaningful and appropriate information for school administrators and policymakers to support evaluation and enhancement of curriculum and programs, and
- Providing appropriate information to support well-designed and meaningful federal and state accountability determinations.

This chapter presents the Task Force’s recommendations for a Comprehensive Wyoming Assessment System in a holistic manner: by painting a verbal picture of how such an assessment system might look. We describe a vision of an assessment system that blends high-quality state assessment results with results from local assessments to support the multiple purposes described above. A comprehensive listing of specific recommendations is given in Chapter 5.

### **Context**

Wyoming, in response to the Campbell decisions among other reasons, has a long history of supporting and requiring district assessment systems to document that all Wyoming students have had an opportunity to learn the “basket of goods,” a term used to describe the knowledge and skills in nine content areas. Any proposed comprehensive assessment system must address how the state will monitor student learning of this basket of goods.

Wyoming’s educational system, in spite of the centralized funding model, is notably local control. Therefore, the Assessment Task Force is recommending an approach to assessment that supports the multitude of uses described above, but that strongly values and improves the quality of locally-generated information.

### **Proposed Wyoming Assessment System**

The Wyoming Assessment Task Force recommends designing and implementing an assessment system that relies on local assessment results to provide rich information to support instructional and evaluative decisions, while relying on state summative assessments to support accountability decisions. This is done by focusing on improving assessment practice, quality of data produced, and the data generated by four main components:

1. **State** course-end or year-end **summative assessments in grades 3-10** designed to support state school (and perhaps district) accountability decisions, serve program evaluation needs at local, regional, and state levels, and to audit local assessment results.
2. **State-supported**, district **interim assessments** are designed to provide checks on student performance a few times during the school year. The interim assessments paid for by the state were useful to many districts, but the Task Force was concerned about the lack of coherence between the interim and summative assessment systems. Therefore, the Task Force recommends that as part of the contract for state summative assessment, the state also contracts for an interim assessment tied to the summative assessment that local districts may choose to use. Districts may choose to select another interim assessment, but if it does so, costs are the responsibility of the district.
3. The **district assessment system** used to document students' opportunity to learn the "basket of goods" can take many forms ranging from district-selected or -created end-of-course summative to assessments to end-of-unit or similar assessments that are aggregated over the course of a year to produce determinations of student performance in specific courses/grades.
4. **Classroom formative assessment** practices designed and implemented by teachers to inform moment-to-moment monitoring of student learning, inform immediate adjustment of instruction, and engage students in monitoring and furthering their own learning.

The common targets of the assessment system are the Wyoming content standards in each of the nine required content areas. One of the ways in which coherence is designed into the comprehensive assessment system is to ensure that all components of the system are tied to specific standards in each content area.

### **Classroom Formative Assessment**

The Wyoming Assessment Task Force acknowledged the critical importance of classroom formative assessment practices for improving student learning, but emphatically argued that other than briefly discussing formative assessment in this report, the Task Force should remain relatively silent on recommendations related to formative assessment. Task Force members noted formative assessment is the purview of districts (actually, schools and classrooms) and for the most part should not be part of the "state" comprehensive assessment system. The Task Force, however, acknowledged that it would make sense for the state and districts (perhaps organized regionally) to partner in providing high-quality professional development to support increasing local formative assessment practices.

### **District Assessment System**

District summative assessments (e.g. unit-end, mid-term, marking-period assessments) are closely tied to both units of instruction in the classroom and longer term learning outcomes. District summative assessments are aligned directly to the Wyoming content standards in the given grade



level, course, or credit. In fact, district summative assessments may consist of the combination of multiple unit-based assessments if the assessments can be demonstrated to have adequate quality and adequate alignment to the applicable Wyoming content standards (e.g., the grade level, course, or credit). The Task Force noted that district assessments play multiple roles, contingent upon the intended uses in the respective districts. For example, district assessments may serve both an auditing function for individual teachers' understanding of their students' learning, and a signaling function of the kinds of knowledge and skill that should be prioritized in daily instruction and classroom assessment.

Standardization across districts is not necessarily the focus. Rather, the districts use the assessments, tied to local curriculum, but aligned to state standards, to support local decisions around improving instruction and programs. The districts also use these assessments to build a credible set of evidence for each student to document the degree to which he or she has met the required standards.

The Wyoming Assessment Task Force recommends that district assessment results should NOT be used as part of school accountability determinations. The Task Force acknowledged that such a stance may relegate district assessment results to a lower status than the state assessment, but at the current time, Task Force members noted that districts were wrestling with the tension between quality and practicality. In other words, district leaders understand that if district assessment results were used for school accountability purposes, they would need to be higher quality than they are now. At the same time, Task Force members were concerned that it was not practically feasible to dramatically improve the quality of district assessments in the short term so they could be used as accountability indicators. That said, the Task Force recognized the need for improving the quality of district assessment and offered several recommendations to address this need.

While Wyoming districts have been responsible for developing local assessment systems for a long time, there has been little review of the technical quality of such assessment systems. Therefore, the Task Force supported moving to a system that focused on improving the quality of district assessment system through the use of multiple strategies including increasing the assessment expertise of those tasked with reviewing district assessments as part of district accreditation processes and to foster local assessment expertise through state support of district assessment consortia, whether regionally or statewide.

### **Interim Assessments**

The Wyoming State Legislature has required and paid for the implementation of a common interim assessment for all Wyoming school districts. The State supported two administrations of the interim assessment—fall and spring—but many districts paid to support the winter administration of the interim assessment. While many district leaders found value in the commercially-selected interim assessment products, using them for a variety of purposes including documenting within-year growth and identifying students in need of remediation, there was some concern expressed by the Task Force members that it was difficult to coherently interpret the results of the interim assessments in light of the summative assessment expectations because the two assessments were designed to measure different learning targets and in different ways.

The major recommendation coming from the Wyoming Assessment Task Force regarding the interim assessment was that the State should require the development of an interim assessment system based on the same assessment framework and tied to the same learning targets as the state

required summative assessment. Districts may choose to adopt the state-provided interim assessments and would have local control over how they would administer and use the interim assessment results. Districts would have the option of purchasing/developing an interim assessment system not tied to the state assessment system, but such districts would be responsible for the costs of the interim assessment. The only exception to this flexibility is that districts participating in the Bridges Summer School program would be required to administer the fall version of the state-supported interim assessment in order to help evaluate the data of the Bridges program.

[NOTE: We need more discussion about specific RFP-type recommendations for the interim assessment design.]

### **State Summative Assessment**

State, standards-based summative assessments will be required in English language arts (including writing) and mathematics in grades 3-10 as well as a science assessment in grades 5, 8, and 10. These assessments must be designed to fully measure the Wyoming content standards and to assess whether students are on track towards college and career ready outcomes. The Wyoming Assessment Task Force recommends that the grade 10 state summative assessment should count as part of the Hathaway scholarship determinations to explicitly tie the scholarship to the official Wyoming content standards and to assure adequate student motivation on the grade 10 test.

The Task Force acknowledges that it is not appropriate to include all of the high school Wyoming standards on a test given in grade 10, because students still have at least two more years of school remaining. Therefore, Task Force recommends having the Wyoming Department of Education convene a standards review committee to determine which of the Wyoming high school content standards are eligible for testing by the end of 10<sup>th</sup> grade. Because grades 11 and 12 remain important, district assessment systems are required to cover the Wyoming high school content standards that do not appear on the state summative assessment.

The state summative assessment must comply with state and federal laws and with industry best practices, including consistency with professional standards, use of principled assessment design, and minimizing burden on local districts and students. The Task Force strongly recommends that in content areas where it is possible, the state summative assessment is used in at least one other state (preferably many states). There are two reasons for this: to allow for comparison of Wyoming educational outcomes to other states and to encourage a stable state summative assessment over time. In other words, changes to the state summative assessment should be minimized by requiring negotiation with other states and/or a vendor in order to make changes to the assessment system.

The Task Force also recommends that the state continue to fund in-school administration of a college entrance examination in grade 11. Students may opt out of the college entrance examination, but must do so affirmatively. The provision of an in-school opportunity for college entrance testing (rather than a traditional Saturday administration) and the requirement of an affirmative opt out is intended to maximum number of students taking the college entrance test. This high-school test schedule reserves the spring of grades 11 and 12 for college entrance, International Baccalaureate, and Advanced Placement testing. The lack of state mandated testing in grade 11 and 12 facilitates high schools focusing juniors and seniors on highly individualized pathways incorporating career and technical education and/or college preparation. It also allows for strengthening the connection from

the end of high school through college, specifically the connection between grades 11-12 and post-secondary education or training (i.e., grades 13-14).

In order to improve reporting timelines for use in school improvement and other evaluation activities, the Task Force recommends administering state summative assessments online except in isolated situations with emergent needs for paper and pencil. Safeguards for assuring a successful transition to online testing are described in the next chapter. The Task Force recommends administering the summative tests in a three-week window near, but not at, the end of the school year to maximize the amount of instructional time before the test but also assure return of results in time for summer school improvement activities and to support district program evaluation needs.

The Task Force recommends having the state assessments serve both an auditing function for district assessment results and a signaling function of the kinds of knowledge and skill that should be prioritized in district assessments. Therefore, while state assessments include traditional test questions targeted toward lower-level knowledge and skills described in state standards, they also include tasks requiring deeper levels of thinking as described in the state standards, even if such thinking is difficult to measure in an on-demand standardized assessment.

However, the task force is concerned that including too many performance or other extended-response tasks on the state summative assessment may lead to unacceptable testing times. That said, the Task Force strongly recommends that such questions occupy a visible place in the state assessment system to help signal the types of assessment tasks the state would like to see on classroom and district assessments and to ensure that the state assessments can provide information about student learning of the full depth of the content standards.

To assure that testing time for state summative assessment is reasonable, a yearly limit on state testing is put in place: no more than *one percent* of the Wyoming required instructional hours for any grade can be required for state testing. Because in Wyoming, state tests are not timed, this limit is based on estimated testing time requirements. These estimates are improved in accuracy with experience over time.

### **Supports and Conditions**

To improve fidelity of implementation at the classroom, school, district, and state level, the Task Force recommends ensuring that following conditions and supports are put in place.

1. To facilitate information flow between district and the state, a two-way data exchange is implemented. Flowing from the state to the district, state-level data are transmitted to the same local district electronic systems mentioned above, where teachers and administrators can access individual and aggregate state, local, and classroom data for their students. Flowing from the district to the state are district-level standards-based designations from district summative assessments. These links can be used to audit district-level standards-based designations.
2. District data systems are developed to house a sophisticated electronic portfolio system that enables it to keep track of students' work along with scores for each of the required standards and skills.
3. The state should contract with a vendor with experience in high-quality interim and summative assessment including performance tasks and projects to measure high-level

knowledge and skills. Districts that request assistance in developing or refining local systems may request assistance from the vendor staff on a cost optional basis.

4. Districts should set aside funding for formative assessment coaches to work with school and district teacher teams to improve diagnoses, feedback, and how to use the results to improve instruction and student learning.
5. Finally, the state should contract for an ongoing evaluation of (1) the quality of the state assessment; (2) the degree to which intended short-, mid-, and long-term outcomes are being achieved; (3) the degree to which anticipated and unintended consequences have been observed and minimized (for the unintended, negative consequences); and (4) after three to five years, a summary report including potential improvements to the system to address any issues identified.

### Ensuring a Successful Transition

The Task Force is mindful of painful experiences with online assessment in the past. In order to ensure a successful transition to online state summative assessment and high-quality interim assessment systems, the transition will be implemented over a period of three years. **Allowing the full three years from the time of acting up these recommendations is critical to assuring that the transition is successful.** The first all-online administration of the state summative assessment will take place in the spring of 2018 and the transition must be smooth. The Task Force recommends a comprehensive set of safeguards to assure a smooth transition, as follows:

1. Schools and districts will be notified immediately that they must be ready for online assessment in spring of 2018.
2. The state will immediately contract for a high-quality comprehensive technology infrastructure audit for the state as a whole and for every school and district. The state audit will at a minimum cover adequacy of the state internet backbone. District audits will at a minimum cover adequacy of available bandwidth, stability of connections to the state backbone and/or other networks. School audits will at a minimum cover adequacy of available bandwidth, stability of connections to district/state systems, adequacy of wireless school network capacity, adequacy of the number of devices capable of administering the assessment, and the adequacy of the operating systems used on those devices.
3. The state contractor will work with each school district to assist in performing the audit (including fully conducting the audit if necessary) to assure a consistent application across all districts.
4. The state contractor will produce a report for the state as a whole (including a summary of district and school reports), each district (including a summary of each school report), and each district. The report will identify specific gaps in technology infrastructure in each report and identify minimum actions that must be taken to close those gaps.
5. The Wyoming State Superintendent will issue a directive that districts must prioritize the use of existing appropriations to assure that gaps are filled.
6. All appropriate state agencies that will support school technology infrastructure should pledge their support for preparing all schools for online assessment by spring 2019 and clearly describe what forms their support will take.
7. At least ten months in advance of the first online administration, all schools, districts, and the state contractor will conduct a simultaneous load test simulating all of Wyoming's students logging on and taking the test simultaneously to attempt to "break" the system. Any

breaks or near breaks in the system as a result of the load test will be used to increase capacity in any areas necessary before the first administration.

8. A paper and pencil option must be available to address isolated emergent needs that cannot be resolved in a reasonable amount of time to allow for online testing.
9. Schools should have reasonable flexibility on scheduling testing within the test window to accommodate the use of online assessment with a limited number of devices (e.g., the length and number of test sessions for each student).
10. It will be communicated often to both parents and educators that prior to taking assessments online, students should be provided with adequate experience in the classroom using devices they will take the test on. This should include at a minimum specific focus on navigating a screen and keyboarding. The Department of Education should gather a workgroup of educators to develop guidelines for providing adequate experience.

The next chapter offers more detailed recommendations that should assist the Department of Education in developing a request for Proposals (RFP) and evaluating bids for a new Wyoming Comprehensive Assessment System.

## CHAPTER 5: DETAILED RECOMMENDATIONS

The following detailed recommendations were put forward by the Task Force with the understanding that they should generally not be written into statute or rule, except for the modification of existing statute and rule currently prohibiting the enactment of these recommendations. These recommendations should instead be embodied in a Request for Proposals (RFP) to be issued so that vendors can bid on providing the services required to implement the system. This understanding is important in that it allow for minor adjustments as necessary. However, it would be reasonable to require reasonable compliance with these recommendations where it is feasible to do so and where an unanticipated compelling reason to choose a different course does not arise.

### State Summative Assessment

#### Recommendations to Support the Intended Outcomes

- Eliminate the grade 11 state required summative assessment.
- To maintain the benefits of a statewide administration of a college entrance examination (but without inappropriate use of the data), provide for a state-paid opportunity for every grade 11 student to take a college entrance examination (but do not require it). Based on prior experience when the college entrance examination was optional, more than 90% of students took the assessment. To maximize the number of students taking the test (without requiring students who object to taking the test to do so) put the following requirements in place:
  - Administer the college entrance assessment in school on a regular school day.
  - Require students to actively opt out of the assessment by a reasonable deadline if they desire not to take the assessment.
- In order to clearly identify what content is eligible to appear on the grade 10 test, the Department of Education should facilitate a standards review committee with the charge of specifying which of the Wyoming content standards are expected to be taught and learned by end of grade 10. The committee should be composed of K-12 content specialists, district curriculum directors, and higher ed content specialists.
- Require state summative assessment in grades 3-10 (adding grades 9 and moving the grade 11 assessment down to grade 10). Science should still be tested by grade span rather than in every grade.
- Fully align the grade 9 and 10 assessments to Wyoming's state content standards.
- To ensure both (1) student motivation on the grade 10 test, and (2) alignment of the Hathaway scholarship criteria with the official Wyoming content standards, include the grade 10 assessment in the criteria for Hathaway scholarship eligibility, with opportunities to retest in grades 11 and 12.
- The Department of Education should be provided with funding for a contract to conduct studies to develop predictive relationships between the grade 9 and 10 assessments and the college readiness assessment.
- The Department of Education should be provided with funding for a contract to provide students with detailed information about their career/college interests and development of individualized high-school pathways.

## Foundational Principles

- To avoid loss of negotiating power and ability to meet Wyoming’s needs, assessment products should not be named in statute, rule, or policy. Nor should statute, rule, or policy so tightly define requirements that only one product is qualified.
- The technical quality of the assessment should be well-documented according to research and/or best practices as referenced by some or all of the following:
  - Principled assessment design (e.g., *Evidence Centered Design*<sup>7</sup>, *Knowing What Students Know*<sup>8</sup>)
  - Universal Design for Learning<sup>9</sup>
  - The AERA/APA/NCME *Standards*<sup>10</sup>
  - CCSSO/ATP *Best Practices for Statewide Assessment*<sup>11</sup>
  - Applicable state and federal law and regulation
  - Federal peer review requirements

## Testing Time

- Actual testing time for state-required summative assessment is not to exceed 1% of the required instructional hours for a given grade level (based on Chapter 22 of Wyoming Department of Education rules, this is a maximum of 9, 10.5, and 11 hours of testing time for elementary, middle, and high school, respectively).
- “Actual testing time” means the time that students are actually responding to assessment tasks (not additional time used for test preparation, breaks, gathering students, logging students, or reading test instructions). Because Wyoming state assessments are not timed, “actual testing time” is based on estimated testing time needed for 85% of students to complete the test. These estimates should be updated annually based on actual test administration.
- The Department of Education should work with a committee of stakeholder responsible for district and school calendaring to finalized testing windows and to assure that districts have sufficient flexibility to address calendaring needs in scheduling assessment.

## Moving Assessment Online

- Test administration must be fully online to expedite return of assessment results and the use of data in school improvement activities and evaluation and adjustment of instructional approaches, curriculum, and programming
- Given Wyoming’s problematic history with online assessment, the transition must be smooth. Several safeguards will be put in place to assure a smooth transition, which are listed below. The most important of these is that the new assessment system should be developed and implemented over three years. If these recommendations are acted upon quickly, a new assessment system could be in place by spring of 2018.
- Schools and districts will be notified immediately that they must be ready for online assessment in spring of 2018.

---

<sup>7</sup> Need citation.

<sup>8</sup> Pellegrino, Chudowsky, & Glaser (2001).

<sup>9</sup> Need citation.

<sup>10</sup> Need citation.

<sup>11</sup> Need citation.

- The state will immediately contract for a high-quality comprehensive technology infrastructure audit for the state as a whole and for every school and district. The state audit will at a minimum cover adequacy of the state internet backbone. District audits will at a minimum cover adequacy of available bandwidth, stability of connections to the state backbone and/or other networks. School audits will at a minimum cover adequacy of available bandwidth, stability of connections to district/state systems, adequacy of wireless school network capacity, adequacy of the number of devices capable of administering the assessment, and the adequacy of the operating systems used on those devices.
- The state contractor will work with each school district to assist in performing the audit (including fully conducting the audit if necessary) to assure a consistent application across all districts.
- The state contractor will produce a report for the state as a whole (including a summary of district and school reports), each district (including a summary of each school report), and each district. The report will identify specific gaps in technology infrastructure in each report and identify minimum actions that must be taken to close those gaps.
- The Wyoming State Superintendent will issue a directive that districts must prioritize the use of existing appropriations to assure that gaps are filled
- All appropriate state agencies that will support school technology infrastructure should pledge their support for preparing all schools for online assessment by spring 2019 and clearly describe what forms their support will take.
- At least ten months in advance of the first online administration, all schools, districts, and the state contractor will conduct a simultaneous load test simulating all of Wyoming's students logging on and taking the test simultaneously to attempt to "break" the system. Any breaks or near breaks in the system as a result of the load test will be used to increase capacity in any areas necessary before the first administration.
- A paper and pencil option must be available to address isolated emergent needs that cannot be resolved in a reasonable amount of time to allow for online testing.
- Schools should have reasonable flexibility on scheduling testing within the test window to accommodate the use of online assessment with a limited number of devices (e.g., the length and number of test sessions for each student).
- It will be communicated often to both parents and educators that prior to taking assessments online, students should be provided with adequate experience in the classroom using devices they will take the test on. This should include at a minimum specific focus on navigating a screen and keyboarding. The Department of Education should gather a workgroup of educators to develop guidelines for providing adequate experience.

### **Test Timing and Test Windows**

- State testing should occur during a three- to four-week testing window which is the same for every grade, with the last allowable testing day being in the first half of May.
- All aggregate reports (other than statewide aggregate reports) should be available by August 1 to facilitate school improvement activities (with consideration that in the first year of any new program, reports are likely to be delayed).
- The Department of Education should survey districts to set for each year a first allowable and last allowable testing day consistent with the previous recommendation. If possible, start and end times should be later to maximize instruction before assessment, but should also



consider typical year-end school activities and the time needed to return data to schools in time for school improvement activities.

### **Claims that Must Be Supported for Individual Students**

- How each student achieves relative to Wyoming standards, including more difficult to measure, high-level knowledge and skills.
- How each student achieves in producing high-quality writing (requires at least two extended written responses per student to support this claim).
- How each student gained in learning relative to peers.
- Student achievement and growth scores are accurate across the range of student achievement, meaning that:
  - Scores are generally free of floor or ceiling effects.
  - Scores support claims about whether novice, typical, and advanced students are being well educated.

### **Claims that Must Be Supported for Classrooms, Schools, Districts, and the State**

- The magnitude of achievement and growth gaps key demographic groups (e.g., sex, race/ethnicity, economic disadvantage, special education, and English learners)?
- The change in achievement and growth gaps over time.
- The percentage of Wyoming students meeting proficiency targets.
- The percentage of Wyoming students meeting growth targets adequate to remain proficient (for already proficient students) or to achieve proficiency (for not yet proficient students) within a reasonable number of years.
- Produces valid and reliable group reports (at the class, school, district, and state level) on strengths and weakness in both proficiency and growth in a small number of sub-areas of each content area. This supports school improvement activities, post hoc evaluation of instructional practices, curriculum, and programming, and high level policies. This could be accomplished using green/yellow/red light reports that show for each group the sub-areas in which a group's achievement is better than, similar to, or worse than its overall content area achievement.

### **Reporting**

- Reports must be designed to meet the needs of the following four groups of stakeholders with similar interests:
  1. Students and parents
  2. Teachers
  3. School and district leadership teams
  4. Business community, media, State School Board, State Superintendent, Joint Legislative Education Committee, Legislature at large, Governor, and general public
- Individual student reports must be designed with stakeholder groups 1 and 2 in mind
- Aggregate reports (e.g., classroom and school reports) showing individual student data must be designed with stakeholder groups 2 and 3 in mind.
- Aggregate report showing group summary data must be designed with all four groups of stakeholders in mind.

- Unless it is possible to adequately serve the needs of multiple stakeholder groups with a single report format, each report should be developed with a format specific to each audience.
- The format and elements of each report should be determined by conducting focus groups and/or multiple rounds of workshopping, with a focus on the following for each report element:
  - What is the “so-what” message appropriate for the audience.
  - Clarity and transparency of the “so-what” message.
  - Approaches to minimize probable misinterpretations.
  - Consistency with AERA/APA/NCME Standards for score reporting.
- The reporting system should allow for teachers to receive dynamic individual reports for just their current students, and aggregate reports for their current and past students.
- The reporting system should allow for each audience to obtain the desired information using intuitive navigation and assistance in finding reports to answer specific questions. Report users should be able to retrieve data to answer their questions with a minimum number of clicks through guided selection of options. Where access to data is appropriate, report users should be able to easily retrieve data about achievement and growth for individual students and demographic groups at the student, classroom, school, district, and state level; with simple navigation between levels.

### **Avoiding a Custom Wyoming Assessment**

- Each content area test must be used in at least one other state (preferably several other states) for the following reasons:
  - Provide stability by requiring changes to the assessment to be negotiated with at least one other state and/or vendor.
  - Facilitate comparison of results from the Wyoming assessment to results from other states.
- To maximize market competition and the ability to meet Wyoming’s needs and to avoid loss of negotiating power, recommendations in this section should be required only where there are at least two options available.

### **Wyoming Educator Participation in Ongoing Development**

- Although avoiding a custom Wyoming assessment means that development of the assessment will already be completed, it is desirable that Wyoming educators have the opportunity to be involved in ongoing development of the assessment. However, to avoid restricting market competition, the recommendations in this section are different in that they would not be considered a “requirement” for a successful bid but a value-added benefit that vendors would be encouraged to consider:
- Wyoming educators have substantive say in ongoing development activities including item development, item review, rangefinding, and other development activities.
- Wyoming educators have the opportunity to review test questions for specific Wyoming sensitivities.
- If there are alternative test questions available to replace those flagged as problematic by Wyoming educators, WDE is able to replace the flagged questions.
- Wyoming educators are involved in scoring student responses requiring human scoring for tests completed by Wyoming students

- The Wyoming Department of Education defines and oversees Wyoming educator involvement.

### **Test Security**

- In order to avoid the stress and disruption of test security breaches, the Department of Education must develop a high quality policy document and associated training using industry standards on test security.
- The policy document and training must include clear policies, protocols, and guidelines to comprehensively address test security in all aspects of testing including at least the following areas:
  - Professional development
  - Prevention of test security breaches
  - Detection of test security breaches (including balancing protection for whistleblowers and minimizing the impact of malicious allegations)
  - Investigating potential security breaches
  - Protocols for evaluating evidence to make conclusions
  - Protocols for appeals of conclusions
  - Follow-up activities to a substantiated or suspected security breach
- The Department of Education's test administration vendor must assist with test security to supplement agency capacity in each of the areas listed in the previous recommendation.
- The Department of Education's test administration vendor must document its own security procedures throughout its processes.

### **Program Evaluation**

- The state should contract for an independent summary report evaluating the degree to which the intended outcomes of the state summative assessment have been realized after five years of implementation
- The evaluation should include the following at a minimum:
  - The quality of the state assessment
  - The degree to which intended short-, mid-, and long-term outcomes are being realized
  - The degree to which anticipated/unanticipated/unintended consequences have been observed
  - Should this be an ongoing evaluation, or does this invite instability?
- To monitor for concerns before and after the five-year evaluation, and to make recommendations as needed, the Department of Education should empanel from this point forward a statewide assessment policy advisory group (PAC) that meets at least twice a year. This panel should include teachers, administrators, technology coordinators, and assessment coordinators.

### **Alternate, English Proficiency, and K-3 Literacy Assessments**

The Task Force focused its efforts on designing a coherent assessment system for the general student population in the content areas comprising the basket of goods. The Task Force also recognizes the importance of coherence of its recommendations with alternate assessment for students with significant cognitive disabilities, English proficiency assessment for English language

learners, and early literacy assessment in grades K-3. However, the Task Force also recognizes specialists in special education, English proficiency, and early literacy are needed to make appropriate recommendations for these specialty assessments. Therefore, the Task Force recommends that in each of these three areas, the Department of Education convene committees to review the recommendations for state summative assessment presented in this report. Those committees should make recommendations for those assessments to be coherent with the general content area assessments by determining which of the recommendations in this report are appropriate for those assessments, which are inappropriate, which need to be modified, and to identify any additional recommendations that may be needed.

### **Interim Assessment**

The Task Force recommends that the state support an interim assessment system to encourage consistency across the state. The use of interim assessments should be governed by the following principles:

- Districts would not be required to adopt the state-provided interim assessment, but districts choosing to administer a different interim assessment system would do so at their own expense.
- The interim assessment supported by the state should be coherently tied to the state summative assessment system in terms of learning targets, format, and design to assure a consistent experience for educators across both state summative and interim assessment.
- To achieve competitive pricing and coherence, the interim assessment would be procured as part of the summative assessment RFP process.
- Interim assessments should provide a check on the big ideas associated with the grade level learning targets to provide an outside audit of the district assessment results.
- In two cases districts may be required to adopt the state-supported interim assessment at the discretion of the Department of Education:
  - The district accepts funding to implement a Summer Bridges program, in which case at least the fall interim assessment must be administered
  - The district has low scores on WAEA (Wyoming's accountability system)

### **District Assessment System**

As the major issues identified with district assessment systems are uneven quality and uneven coherence with state assessment, several recommendations address quality control and information flow.

- To facilitate information flow between district and the state, a two-way data exchange should be implemented. Flowing from the state to the district, state-level data are transmitted to local district electronic systems, where teachers and administrators can access individual and aggregate state, local, and classroom data for their students. Flowing from the district to the state are district-level standards-based designations from district summative assessments.
- These links can be used to audit district-level standards-based designations and identify districts with local assessment systems that may need improvement.

- District data systems should be developed to house a sophisticated electronic portfolio system that enables it to keep track of students' work along with scores for each of the required standards and skills to document learning of the basket of goods.
- The state should contract with a vendor with experience in high-quality interim and summative assessment including performance tasks and projects to measure high-level knowledge and skills. This vendor should fill two roles: (1) provide district and school personell with statewide professional development in developing high-quality interim and summative assessments, and (2) for districts that request assistance in developing or refining local systems, provide that assistance on a cost optional basis.
- To improve quality and assure consistency of reviews, the state should contact with one or more qualified professionals to perform audits of district assessment systems as a part of the accreditation process.
- The state should should incentivize and/or support collaborative efforts among districts to improve task and assessment quality. This could be modeled after the WY BOE Assessment Activities Consortium. This could include hosting for educator access intact assessments, banks of high-quality tasks and test questions, and appropriate professional development on using the resources.

Because considerable improvements in district assessment systems would be required to support high-stakes use, the workgroup recommends NOT using the district assessment results as an indicator in WAEA at this time.

### **Classroom Formative Assessment**

The Wyoming Assessment Task Force acknowledged the critical importance of classroom formative assessment practices for improving student learning, but emphatically argued that other than briefly discussing formative assessment in this report, the Task Force should remain relatively silent on recommendations related to formative assessment. Task Force members noted formative assessment is the purview of districts (actually, schools and classrooms) and for the most part should not be part of the “state” comprehensive assessment system. The Task Force, however, acknowledged that it would make sense for the state and districts (perhaps organized regionally) to partner in providing high-quality professional development to support high-quality local formative assessment practices.

## CHAPTER 6: RECOMMENDATIONS FOR POLICY COHERENCE

Placeholder

DRAFT

## REFERENCES

- Braun, H. (Ed.). (in press). *Meeting the Challenges to Measurement in an Era of Accountability*. Washington, DC: National Council on Measurement in Education.
- Coladarci, T. (2002). Is it a house...or a pile of bricks? Important features of a local assessment system. *Phi Delta Kappan*, 83(10), 772-774.
- Michigan Department of Education. (2013). *Report on Options for Assessments Aligned with the Common Core State Standards*. Retrieved June 20, 2015, from [http://www.michigan.gov/documents/mde/Common\\_Core\\_Assessment\\_Option\\_Report\\_441322\\_7.pdf](http://www.michigan.gov/documents/mde/Common_Core_Assessment_Option_Report_441322_7.pdf)
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC. Retrieved September 4, 2015, from [http://www.nap.edu/openbook.php?record\\_id=10019&page=R1](http://www.nap.edu/openbook.php?record_id=10019&page=R1)
- Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief*. Retrieved June 20, 2015, from [http://www.nciea.org/publication\\_PDFs/PolicyBriefFINAL.pdf](http://www.nciea.org/publication_PDFs/PolicyBriefFINAL.pdf)
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
- Wiley, E. C. (2008). *Formative Assessment: Examples of Practice*. Retrieved August 11, 2015, from [http://ccsso.org/documents/2008/formative\\_assessment\\_examples\\_2008.pdf](http://ccsso.org/documents/2008/formative_assessment_examples_2008.pdf)

## **APPENDIX A: VIGNETTES OF FORMATIVE ASSESSMENT PRACTICES**

The following four vignettes describe formative assessment practices of four different teachers to allow educators to better understand what is meant by “formative assessment.”

### **High School – Chemistry Mid-Period Check In**

As part of instructional planning, a high school chemistry teacher develops both true and false statements related to a micro-unit covering a half hour in high school chemistry. Statements were strategically developed to assess whether students hold anticipated misconceptions. Following the micro-unit, students show thumbs up, thumbs down, or thumbs to the side to indicate whether each statement is true, false, or they don’t know. Based on the prevalence of thumbs down and to the side, the teacher may select one of at least four options:

1. Reteach that micro-unit using a different instructional plan the next day.
2. Use pre-planned strategies to address a small number of misconceptions.
3. Strategically group students who put thumbs down or to the side with confident students to discuss their conclusions and monitor group discussions.
4. Work briefly with a one or two students needing additional assistance while the rest of the class engages in the next activity.

### **Middle School – English End of Period Check In**

At the beginning of a seventh grade English class period, a middle school English teacher shares with her students what the three learning targets are for the day. At the end of the period, she asks each student to fill out and hand in a slip confidentially rating their attainment of each learning target in one of the following four categories:

1. I can teach this.
2. I can do this on my own.
3. I need some help with this.
4. I don’t get this at all.

The teacher adjusts the next day’s lesson plan by creating a simple task asking small groups of students to practice a learning target on which about half the students felt confident. The small groups are strategically selected to include students that are both confident and not confident with the learning target. She also reviews with the entire class another learning target on which few students felt confident. To do so, she asks two students to explain their approach on a specific problem. After gauging current understanding, she decides whether to instruct on that learning target again using a different strategy and different examples than the previous day.

### **Elementary School – Monitoring Development of Mathematical Understanding**

After a successful unit on simple two-digit addition (without regrouping), an elementary school teacher wants students to learn both a regrouping algorithm and why the algorithm works. He demonstrates to his students that their current knowledge and skills are inadequate to accurately deal with two-digit addition requiring regrouping. He does this by assigning small groups of students to solve a problem either using the addition algorithm they already know or by using counting objects.



In a subsequent whole-class discussion, the teacher highlights the conflicting answers and asks his students to think about how place value place might explain why the groups got different answers. He then asks each small group to work on developing its own solution to the problem. After visiting and probing each group to survey current understanding and developing strategies, he asks strategically chosen groups to share their developing solutions, and builds post-activity instruction on the regrouping algorithm around them.

### **High School – English Capstone Project**

As a capstone project for a unit on persuasive writing, a high-school English teacher assigns her students to individually write a persuasive essay incorporating each of the unit learning targets. Each student is to:

- Choose a position on a controversial topic important to him,
- Identify reliable resources for information on his position and a contrary position commonly taken on the topic,
- Summarize the arguments for both positions,
- Use the logical devices taught in the unit to argue for his position,
- Use logical tools to argue the logical superior of his position, and
- Incorporate work in all five previous steps into a coherent persuasive essay.

The teacher divides the capstone project into four subunits (with associated assignments):

1. Choosing a topic, a personal position, an opposing position, and identifying reliable resources;
2. Summarizing arguments for at least two positions on the topic;
3. Arguing for the personal position and against an opposing position on a logical basis;
4. Incorporating into a complete and coherent persuasive essay.

Along with other formative practices, the teacher spends class time making each sub-unit's learning targets explicit and instructing on them. She also uses class time on the day each assignment is due to have students peer-review each other's work, focusing on the learning targets and working on revisions. As assignments are turned in, the teacher provides formative feedback based on the learning target rather than grading each assignment. Only after providing at least one round of formative feedback on each assignment does the teacher grade the final product. She does this to ensure that the formative feedback fulfills its purpose and her evaluation of each student's performance represents what was learned by the end of the unit.

## APPENDIX B: ONE-PAGE SUMMARY OF FORMATIVE, INTERIM, AND SUMMATIVE ASSESSMENT

	Formative Assessment	Interim Assessment	Summative Assessment
Characteristics	<ul style="list-style-type: none"> <li>Facilitate effective instruction (does not pause instruction)</li> <li>Learning goals and criteria are clear to students</li> <li>Students self-/peer-monitor progress toward learning goals</li> <li>Students and teachers receive frequent feedback</li> <li>Jointly controlled by each teacher and her students</li> <li>Covers a micro unit of instruction</li> <li>Very frequent (e.g., multiple times per period)</li> <li>Tailored to a set of students and an instructional plan</li> <li>Might be comparable for a classroom, but not beyond</li> <li><i>Not a product (e.g., quiz, test, bank of questions/ tests)</i></li> </ul>	<ul style="list-style-type: none"> <li>Pauses instruction for evaluation</li> <li>Controlled solely by a teacher, school, district, or state (or by a consortium of teachers, schools...)</li> <li>Covers a mid-sized unit of instruction</li> <li>Somewhat frequent (e.g., weekly to quarterly)</li> <li>Administered before and/or after a mid-sized unit</li> <li>Based on who controls assessment, results may be comparable across students, teachers, schools, districts, and/or states</li> <li>A product</li> </ul>	<ul style="list-style-type: none"> <li>Pauses instruction for evaluation</li> <li>Controlled solely by a teacher, school, district, or state (or by a consortium of teachers, schools...)</li> <li>Covers a macro unit of instruction (e.g., semester, course, credit, grade)</li> <li>Infrequent (e.g., yearly, finals week)</li> <li>Administered after completing a macro unit</li> <li>Based on who controls assessment ,results may be comparable across students,..., and/or states</li> <li>A product</li> </ul>
Uses	<ul style="list-style-type: none"> <li>Engage students in learning/metacognition through frequent feedback and self-/peer-evaluation</li> <li>Monitor moment-to-moment student learning</li> <li>Diagnose individual students' immediate instructional needs</li> <li>Diagnose immediate group instructional needs</li> <li>Immediately adjust instruction</li> <li>Differentiate instruction</li> <li>Self-evaluate micro-unit instructional effectiveness</li> <li><i>Student results from formative assessment are not appropriate for use in grading or accountability; however, ratings of the quality of formative assessment practice may be appropriate for use in accountability</i></li> </ul>	<ul style="list-style-type: none"> <li>Evaluate achievement after a mid-sized unit</li> <li>Monitor progress within a macro-unit (e.g., semester, course, credit, grade)</li> <li>Corroborate formative assessment</li> <li>Pre-test to tailor unit instructional plans for the group and individual students</li> <li>Identify post-unit remedial needs</li> <li>Mid-course self-evaluation and adjustment of teacher classroom practices</li> <li>Mid-course evaluation and adjustment of school and district policies and programs</li> <li>Predict performance on summative assessment</li> <li>Grading (and possibly accountability)</li> </ul>	<ul style="list-style-type: none"> <li>Evaluate achievement after a macro unit</li> <li>Monitor progress across multiple macro-units</li> <li>Corroborate interim assessment</li> <li>Evaluate readiness for the next macro unit</li> <li>After-the-fact evaluation/adjustment of broad instructional practices by individual teachers and of curriculum/programming policies by administrators</li> <li>Predict later student outcomes</li> <li>Grading and accountability</li> </ul>
Examples	<ul style="list-style-type: none"> <li>Following a micro-unit, students show thumbs up/thumbs down to indicate whether statements developed around anticipated misconceptions are true. Based on prevalence of misconceptions, the teacher reteaches parts of his lesson using a different instructional strategy, strategically groups students to discuss their conclusions, or works briefly with one or two students.</li> <li>At the end of class, students hand in a slip confidentially rating their attainment of each learning target as: (1) <i>I can teach this</i>, (2) <i>I can do this on my own</i>, (3) <i>I need some help with this</i>, or (4) <i>I don't get this at all</i>. The teacher adjusts her next-day group assignments and planned activities accordingly.</li> </ul>	<ul style="list-style-type: none"> <li>Classroom unit quizzes and homework</li> <li>Individual and group unit projects</li> <li>Pre-unit exams of unit pre-requisites</li> <li>Pre-unit exams of unit content</li> <li>End of unit exams</li> <li>Mid-term exams</li> <li>Marking period exams not covering a full macro-unit</li> <li>Quarterly assessments</li> <li>District placement tests</li> </ul>	<ul style="list-style-type: none"> <li>Classroom final exams, projects, and papers</li> <li>School or district final exams, projects, or papers</li> <li>District/state assessments for testing out of a credit</li> <li>District graduation/diploma-endorsement tests</li> <li>Typical state accountability tests</li> <li>High school equivalency tests</li> <li>District graduation tests</li> <li>College admission tests</li> </ul>

## APPENDIX C: DETAILED HIGHEST PRIORITY USES AND CHARACTERISTICS

The Task Force's highest priority uses and characteristics are presented in detail in Table B1 below. These uses and characteristics were evaluated by the facilitators using the definitions and appropriate uses of formative, interim, and summative assessments discussed in Chapter I. The evaluation also incorporates differences between classroom-, district-, and state-owned assessments to show the complexity of an assessment system that would be needed to fulfill all of the Task Force's highest priority uses and characteristics. This evaluation is reflected in additional elements added to Table B1. Those elements identify whether each type and level of assessment has full, some, minimal, or no applicability to the use or characteristic in each row. In addition, in each row the applicability of the various types and levels of assessment to each use or characteristic is briefly explained.

Table B1. Task Force Highest Priority Uses and Characteristics.

Total <sup>1</sup> Score	Number of Votes by Priority			Desired Uses and Characteristics of Wyoming Assessment	Applicability <sup>2</sup>					
					Type			Level		
	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>		Formative	Interim	Summative	Classroom	District	State
38	10	3	2	Provide information to parents, students, and educators regarding individual student achievement and growth within and across years, including readiness for the next level in a student's K-12 progression - Classroom formative: continuous achievement/growth/readiness data on micro-units - Classroom/district/state interim: periodic achievement/growth/readiness data on mid-sized units - Classroom/district/state summative: yearly achievement/growth/readiness data on macro-units	●	●	●	●	●	●
27	6	4	1	Provide feedback on progress toward standards to inform instruction on more than a yearly basis - Classroom formative: continuous achievement and progress data inform daily instruction - Classroom/district/state interim: periodic unit achievement & progress data informs remediation - District/state summative: interim results might be rolled up for summative determinations	●	●	●	●	●	●
16	0	5	6	Allow for comparisons within the state and across states - State interim: provides within-state comparability if adopted statewide - State summative: provides within-state comparability - State interim/summative: provides cross-state comparability if a multi-state assessment is used	○	●	●	○	○	●
13	2	2	3	Provide reliable and valid data to evaluate program/curriculum effectiveness and alignment to standards - District/state interim: can provide information to inform within- and between-year evaluations - District/state summative: can provide information to inform between-year evaluations	○	●	●	○	●	●

Total <sup>1</sup> Score	Number of Votes by Priority			Desired Uses and Characteristics of Wyoming Assessment	Applicability <sup>2</sup>					
					Type			Level		
	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>		Formative	Interim	Summative	Classroom	District	State
11	3	1	0	Be student-centered (e.g., student is not a number) - Classroom formative: micro-unit diagnostic data to tailor instruction - Classroom/district/state interim: unit diagnostic data to tailor remediation - Classroom/district/state summative: macro-unit data to inform critical yearly decisions	●	●	●	●	●	●
8	0	3	2	Encourage collaboration and sharing best practices - Classroom formative/interim/summative: foster teacher collaboration on teacher practices - District/state interim/summative: foster teacher collaboration on using non-classroom data - District/state interim/summative: foster educator collaboration on curriculum/programming - Limit use of classroom assessment for evaluation to quality of practices and support for collaboration	●	●	●	●	●	●
7	1	2	0	Continually inform instruction with timely feedback - Classroom formative: continual micro-unit diagnostic data to inform daily instruction - Classroom/district/state interim: periodic unit data to inform post-unit remediation	●	●	○	●	●	●
6	1	1	1	Validly inform decisions about post-secondary education/training - State summative: likely to provide based on ties to post-secondary outcomes (onerous for a district)	○	○	●	○	●	●
2	0	0	2	Consistency over time to facilitate the intended outcomes of assessment in Wyoming - District interim/summative: stable longitudinal data can improve decision making - State interim: stable longitudinal data can improve decision making - State summative: likely to improve decision-making because of school/district accountability uses	○	●	●	○	●	●
				Number of desired uses/characteristics with unique and full applicability	2	0	3	3	0	3
				Number of desired uses/characteristics with full applicability	4	3	5	4	2	5
				Number of desired uses/characteristics with some applicability	1	4	1	1	4	3
				Number of desired uses/characteristics with unlikely applicability	0	1	2	0	2	1
				Number of desired uses/characteristics with no applicability	4	1	1	4	1	0

- Each panelist identified one characteristic as her highest priority, second highest priority, or third highest priority. These were given scores of 3, 2, and 1 respectively. The scores were summed across panelists to give a total score for each desired use/characteristic.
- , ●, ●, and ○ indicate desired uses or characteristics for which the type or level of assessment has full applicability, some applicability, minimal or unlikely applicability, and no applicability, respectively.